**Appendix 9: Description of All Indices**

1. **STUDENT QUESTIONNAIRE**
2. **Simple Indices**
3. **Students’ Age**

The age of a student was calculated as the difference between the year of the test and the year of the birth reported by the student in the student questionnaire.

1. **Parents’ Occupational Status**

Students were asked to answer two open-ended questions about their parents’ occupational status. A sample question was, “What is your mother’s main job?” Students’ responses were coded to four-digit ISCO codes ([International Labour Office, 1990](#_ENREF_4)). The codes were mapped to the international socio-economic index of occupational status, ISEI, ([Ganzeboom, Graaf, Treiman, & Leeuw, 1992](#_ENREF_1); [Ganzeboom & Treiman, 1996](#_ENREF_2)). Based on the results from the mapping, the following three indices were derived: mother’s occupational status (ISEIM), father’s occupational status ISEIF), and higher occupational status of parents (HISEI).

1. **Parents’ Educational Level**

Indices on parental education level were developed by following the ISCED ([OECD, 1999](#_ENREF_5); [UNESCO, 1997](#_ENREF_8)) standards. Students were asked to provide appropriate schooling levels and qualifications of their parents through close-ended questionnaire items. The schooling levels ranged from not having a primary school education to having a higher secondary school education and the qualifications ranged from having a diploma certificate to having a master degree. The schooling levels were coded as follows: ISCED 1 (primary school education); ISCED 2 (lower); ISCED 3B (middle secondary school) and ISCED 3A (upper secondary school education). The qualifications were coded as follows: ISCED 4 (diploma certificate); ISCED 5A (bachelor degree); and ISCED 6 (master degree). An index of the highest education level (PARED) was also derived by using higher parental education.

1. **Language Spoken at Home**

Information about the language used by students in their homes was collected by asking students to select the language they speak in their homes from a list of languages presented in the questionnaire. Students were also given the option of writing their home language if it was not included on the list.

1. **Expected Educational Level**

Information about students’ expected level of education was collected by asking students to select the level of education that they intended to complete. The choice of the education levels ranged from middle secondary school education to university education.

1. **Grade Repetition**

Students were asked to indicate their grade repetitions at primary school, lower secondary school, and middle secondary school on a three-category Likert-type item, with the categories ranging from 1 (never) to 3 (twice or more).

1. **Late Arrival for School**

Students were asked to indicate the frequency of arriving late for the school in a couple of weeks. The frequency ranged from “none” to “five or more times”.

1. **Mathematics Homework**

Students were asked to provide information about mathematics homework in terms of its frequency and duration. The frequency ranged from “never” to “3 or 4 times a week”, while the duration ranged from “fewer than 15 minutes” to “more than 90 minutes”.

1. **Time Spent on Mathematics Activities**

Students were asked to indicate the amount of time that they spent on different teaching and learning activities related to mathematics. A five-category Likert-type item was used to measure the amount of time, with the categories ranging from 1 (no time) to 5 (6 or more hours a week). A sample question was, “How much time do you typically spend per week studying mathematics?” A sample response statement to the question was, “Remedial classes in mathematics at school”.

1. **Out-of-School Time Lessons**

Students were asked to provide information about the types of out-of-school time lessons that they might have taken. A sample question was, “What type of out-of-school-time lessons do you attend currently?” followed by to a sample response statement, “One to one lessons with a teacher who was also a teacher at your school” with “yes” or “no” options.

1. **Reason for Choosing a School**

Students were asked to provide as many reasons as applicable to them for choosing to attend their schools. A sample item was, “Which of the following are reasons why you attend this school?” followed by a sample response, “This is the local school for students who live in this area”.

1. **ICT**

Students were asked about their experience of using computers, with experience ranging from “less than one year” to “more than five years”. Also, students were asked about the places where they used computers, and the frequency of using computers at those places. A five-point Likert-type item was used to measure the frequency of using computers at home, school, and other places; with the frequencies ranging from 1 (almost every day) to 5 (never). Further, students were asked to indicate who taught them most about how to use the Internet and computers. A sample item was, “Who taught you most about how to use computers?” followed by a sample response statement, “my friends”.

1. **Scale Indices and Construct Validation**
2. **Household Possessions**

A dichotomous item with 14 response statements was used to collect data on the availability of 14 different household possessions at students’ homes. A sample item was, “Which of the following are in your home?” with a sample response statement like “A desk to study at”. The following three different indices were developed from the item: computer facilities at home (COMPHOME), cultural possessions (CULTPOSS), and home educational resources (HERDES).

A confirmatory factor analysis of the three-dimensional model of the students’ household possession, using polychoric correlations with weighted least squares estimator, showed a reasonable model fit (RMSEA=0.46, CFI=0.94, NNFI=0.95), confirming the unidimensionality of the measures. The regression coefficients of the observed variables on their underlying constructs were all significant, with the smallest and the largest t-values being 10.02 and 39.44, respectively; showing convergent validity. The composite reliabilities of the measures were reasonably high (COMPHOME=0.87, CULTPOSS=0.64, HERDES=0.74), indicating good validity. The shared variances between COMPHOME and CULTPOSS (0.17) and between COMPHOME and HERDES (0.40) were less than the average variance explained across the measures for each latent construct (COMPHOME=0.70, CULTPOSS=0.44, and HERDRE=0.37), showing discriminant validity. However, the variance in common between CULTPOSS and HERDES (0.72) was more than the average variance explained across their measures (CULTPOSS=0.44 and HERDES=0.37), indicating non-discriminant validity. This is also clear from their correlation (0.85), indicating that cultural possessions and home educational resources implied approximately the same meaning to students.

1. **SES**

An index of students’ SES was derived as factor scores for the first principal component from the following three variables: (a) HISEI, (b) PARED, and (c) FWEALTH (comprising the IRT weighted likelihood estimates of COMPHOME, CULTPOSS, and HERDES). Prior to factor scoring, the variables were standardized with a mean of zero and a standard deviation of one to assign a common metric. A principal components analysis was performed, by using SPSS version 19, on the standardized variables to obtain factor scores for the first principal component. Similar method was used by PISA ([OECD, 2005](#_ENREF_6), [2009](#_ENREF_7)).

1. **Self-Beliefs**

Students’ self-beliefs were expressed in terms of their self-efficacy, self-concept, and anxiety in relation to learning mathematics. Self-efficacy was measured with a four-category Likert-type item of eight response statements, with the categories ranging from 1 (very confident) to 4 (not at all confident). A sample item on self-efficacy asked students to rate their confidence at solving an equation like, “”. Self-concept was measured with a four-category Likert-type item of five response statements, with the categories ranging from 1 (strongly agree) to 4 (strongly disagree). A sample item on self-concept required students to express the degree of their agreement with statements like, “I learn mathematics quickly”. Anxiety was measured with a four-category Likert-type item of five response statements, with the categories ranging from 1 (strongly agree) to 4 (strongly disagree). A sample item on anxiety required students to express the degree of their agreement with a statement like, “I often worry that it will be difficult for me in mathematics classes”.

A confirmatory factor analysis of the item covariances with robust maximum likelihood estimation method was performed on the three-dimensional model of students’ self-beliefs. The model fitted well (RMSEA=0.42, CFI=0.96, NNFI=0.96, and RMR=0.04), indicating the unidimensionality of the items. The regression coefficients of the observed variables on their underlying constructs were all significant, with the smallest and the largest t-values being 12.51 and 30.78, respectively, indicating convergent validity. The composite reliabilities of the measures were reasonably high (self-efficacy=0.75, self-concept=0.74, anxiety=0.73), indicating good validity. The correlations between self-efficacy and anxiety (-0.59), and self-concept and anxiety (-0.94) are indicative of discriminant validity. Self-efficacy and self-concept did not pass the criterion of discriminant validity. The average variance explained by their measures (self-efficacy=0.28 and self-concept=0.37) was less than their shared variance (0.41). However, their correlation (0.64) was indicative of them being different dimensions of self-beliefs.

1. **School Climate**

Students’ attitude towards their schools, their relationship with teachers, and their sense of belongingness to their schools were used as the three dimensions of school climate. Students’ attitude towards their schools was measured with a four-category Likert-type item of four response statements, with the categories ranging form 1 (strongly agree) to 4 (strongly disagree). A sample response statement that measured students’ attitude towards their schools was, “School has done little to prepare me for adult life when I leave school”. Students’ relationship with their teachers was measured with a four-category Likert-type item of five response statements, with the categories ranging from 1 (strongly agree) to 4 (strongly disagree). A sample response statement on students’ relationship with their teachers was, “Students get along well with most teachers”. Students’ sense of belongingness to their school was measured with a four-category Likert-type item of six response statements, with the categories ranging from 1 (strongly agree) to 4 (strongly disagree). A sample response statement on the sense of belongingness was, “I feel like an outsider”.

A confirmatory factor analysis on the item covariances with robust maximum likelihood estimation method of the three-construct model of school climate fitted well (RMSEA=0.05, NNFI=0.91, CFI=0.93, and RMR=0.03), indicating unidimensionality of the items. The composite reliabilities of the latent constructs were good for students’ attitude (0.74), students’ relationship with teachers (.63), and poor for students’ sense of belongingness (0.57). The regression coefficients of the items on their underlying constructs were all significant, with the smallest and the largest t-values being 6.71 and 29.98 respectively, indicating convergent validity. The average variances explained by the measures of students’ attitude (0.45), relationship with teachers (0.26), and sense of belongingness (0.19) were greater than their variances in common (attitude and relationship=0.04 and attitude and belongingness=0.12), showing discriminant validity. The average variance explained by the measures of students’ sense of belongingness (0.19) was less than the variance in common (0.24) between the students’ sense of belongingness and relationship with teachers, showing the lack of discriminant validity of these constructs. Also, their estimated correlation of 0.52 indicates that they conveyed the same meaning to the students.

1. **Classroom Climate**

Classroom climate was measured as a two-dimensional construct, with teacher support and classroom disciplinary climate as the two constructs. Students were asked to rate their perceptions of teacher support and classroom disciplinary climate in their mathematics class on a four-category Likert-type item, with the categories ranging from 1 (every lesson) to 4 (never or hardly ever). The teacher support and the classroom disciplinary climate were measured with five response statements each. A sample response statement that measured the students’ perceptions of teacher support was, “The teacher shows an interest in every student’s learning”. Similarly, a sample response statement on the classroom disciplinary climate was, “Students do not listen to what the teacher says”.

A confirmatory factor analysis on the item covariances with robust maximum likelihood estimation method of the two-factor model of classroom climate showed a good fit (RMSEA=0.05, NNFI=0.91, CFI=0.93, and RMR=0.04), indicating the unidimensionality of the items. The reliabilities of both the constructs were good (teacher support=0.64 and disciplinary situation=0.64). The regression coefficients of the items on their underlying constructs were all significant, with the smallest and the largest t-values being 5.46 and 20.72 respectively, showing the convergent validity of the items. The average variances explained by the measures of teacher support (0.24) and disciplinary situation (0.41), respectively, were greater than the variance in common between them (0.48), indicating discriminant validity.

1. **Motivation**

Intrinsic and extrinsic motivation were used as the factors affecting students’ motivation to learn mathematics. Students were asked to rate their agreement with a four-category Likert-type item, with the categories ranging from 1 (strongly agree) to 4 (strongly disagree). Both intrinsic and extrinsic motivations were measured with four response statements each. A sample response statement on intrinsic motivation was, “I enjoy reading about mathematics”, and a sample response statement on extrinsic motivation was, “I will learn many things in mathematics that will help me get a job”.

A confirmatory factor analysis on the item covariances with robust maximum likelihood estimation method of the two-factor motivation model showed a good fit (RMSEA=0.06, NNFI=0.96, CFI=0.97, and RMR=0.02), indicating the unidimensionality of the items. The reliabilities of the constructs were good (extrinsic motivation=0.73 and intrinsic motivation=0.75). The regression coefficients of the items on their underlying constructs were all significant, with the smallest and the largest t-values being 16.15 and 29.85 respectively, showing the convergent validity. The average variances across the items of extrinsic motivation (0.41) and intrinsic motivation (0.43) were less the variance in common between them (0.62), indicating non-discriminant validity.

1. **Learning Strategies**

Students’ learning strategies were identified in terms of memorisation, elaboration, and meta-cognition. Students were asked to rate their agreement with the response statements related to each of these learning strategies on a four-category Likert-type item, with the categories ranging from 1 (strongly agree) to 4 (strongly disagree). The memorization, the elaboration, and the meta-cognition strategies were measured with four, four, and five response statements respectively. Sample response statements on the memorization, the elaboration, and meta-cognition strategies were, “To learn mathematics, I try to remember every step in a procedure”, “I try to understand new concepts in mathematics by relating them to things I already know”, and “When I study mathematics, I start by working out exactly what I need to learn”, respectively.

A confirmatory factor analysis on the item covariances with robust maximum likelihood estimation method of the three-factor learning strategy model showed a good fit (RMSEA=0.04, NNFI=0.97, CFI=0.97, and RMR=0.02), confirming the unidimensionality of the items. The reliabilities of the constructs were moderate (memorisation=0.55, elaboration=0.66, and meta-cognition=0.64). The regression coefficients of the items on their underlying constructs were all significant, with the smallest and the largest t-values being 13.08 and 23.04 respectively, showing the convergent validity. The average variances across the items of the constructs (memorisation=0.27, elaboration=0.26, and meta-cognition=0.43) were all less than the variances in common between the constructs (memorisation and elaboration=0.94, memorisation and meta-cognition=0.92, and meta-cognition and elaboration=0.85), indicating that the constructs lacked discriminant validity. This was also clear from the high correlation coefficients of the constructs (memorisation and elaboration=0.97, memorisation and meta-cognition=0.96, and meta-cognition and elaboration=0.92).

1. **ICT Experience**

Students’ ICT usage was assessed in terms of the frequency with which they used the Internet and other computer programmes, and their attitudes towards using ICT. Students were asked to indicate how often they used the Internet and other computer programmes on a five-point rating item of 12 sub-questions, with the points ranging from 1 (almost every day) to 5 (never). A sample sub-question was, “How often do you use the Internet to look up information about people, things, or idea?” The students’ attitude towards using ICT was measured with a four-category Likert-type item of four response statements, with the categories ranging from 1 (strongly agree) to 4 (strongly disagree). A sample response statement on the students’ attitude was, “It is very important to me to work with a computer”.

A confirmatory factor analysis on the item covariances of the three-factor model of the ICT usage with robust maximum likelihood estimation method showed a moderate fit (RMSEA=0.05, NNFI=0.89, CFI=0.90, and RMR=0.08), indicating the unidimensionality of the items. The reliabilities of the constructs were moderate (the Internet=0.64, computer programmes=0.63, and attitudes=0.58).The regression coefficients of the items on their underlying constructs were all significant, with the smallest and the largest t-values being 4.46 and 13.76 respectively, showing the convergent validity. The average variances across the items of the latent constructs the Internet (0.24) and computer programmes (0.22) were less than their common variance (0.64), indicating that the constructs did not pass the discriminant validity. On the other hand, the average variances of the latent constructs (the Internet=0.24, computer programmes=0.22, and attitude=0.29) were greater than their variances in common (the Internet and attitude=0.02 and computer programme and attitude=0.003), indicating discriminant validity. However, the correlation coefficients of attitude and the Internet (-0.15) and attitude and computer programmes (-0.05) were very small and negative.

1. **Confidence with ICT Tasks**

Students’ confidence with ICT tasks was assessed in terms of their confidence in performing routine ICT tasks, the Internet tasks, and high-level ICT tasks. Students were asked to indicate how well they could perform an ICT task on a four-category Likert-type item of 23 response statements, with the categories ranging from 1 (I can do this very well by myself) to 4 (I don’t know what this means). A sample response statement was, “Use a spreadsheet to plot a graph”.

A confirmatory factor analysis on the item covariances of the three-factor model of the students’ confidence with ICT tasks showed a moderate fit (RMSEA=0.08, NNFI=0.96, CFI=0.96, and RMR=0.05), showing the unidimensionality of the items. The reliabilities of the constructs (Routine ICT tasks=0.90, Internet tasks=0.90, and high-level ICT tasks=0.85) were very high. The regression coefficients of the items on their underlying constructs were all significant, with the smallest and the largest t-values being 8.88 and 35.50 respectively, indicating the convergent validity of the items. The average variances across the items of the constructs (Routine ICT tasks=0.47, Internet tasks=0.64, and high-level ICT tasks=0.45) were less than the common variances between the constructs (Routine ICT and Internet tasks=0.61, routine and high-level ICT tasks=0.77, and Internet and high-level ICT tasks=0.722), indicating non-discriminant validity.

1. **TEACHER QUESTIONNAIRE**
2. **Simple Indices**
3. **Teachers’ Demographic Profile**

The information about the demographic profile of teachers was collected through two questionnaire items on their age and gender. The teachers’ responses to the item on the gender were coded 1 for male and 2 for female. The item on the age required the teachers to select their age from a range of intervals, with the intervals ranging from “under 25” to “60 plus”.

1. **Teachers’ Educational Attainment**

The information about the teachers’ educational attainment was obtained through two questionnaire items on teachers’ academic and professional qualifications. The teachers were asked to select their academic qualifications from a list of qualifications provided in the questionnaire item. The list contained academic qualifications from as low as a Grade 10 certificate to as high as a master’s degree. The teachers’ responses were coded by using ISCED, with Grade 10 certificate coded to ISCED 2, Grade XII certificate coded to ISCED 3, bachelor’ degree coded to ISCED 4, and master’s degree coded to ISCED 5. Similar to their academic qualifications, the teachers were asked to indicate their professional qualifications. The list of professional qualifications presented in the teacher questionnaire ranged from a bachelor in education to a master in education. A bachelor degree was coded to ISCED 4 and a post-graduate degree was coded to ISCED 5.

1. **Teachers’ Academic Major**

Teachers were asked whether mathematics was their academic major. A dichotomous question was used to get the teachers’ responses.

1. **Teachers’ Teaching Experience**

Teachers were asked to indicate their experience of teaching mathematics in Grade 10. A sample item on teachers’ teaching experience was, “Counting this year, how many years have you taught mathematics in Grade 10?” The response options for this question ranged from “two years of teaching experience” to “more than five years of teaching experience”. Additionally, teachers were asked to indicate their familiarity with students. A sample item on the teachers’ familiarity with their students was, “Which one of the following best describes your familiarity with students in your class?” A sample response to the question was, “I taught all of them in the precious year”.

1. **Teachers’ Use of Time**

Teachers were asked to indicate the number of hours that they spent, in a typical teaching week, on the following areas: teaching, planning, or preparation of lessons, and administrative duties. A sample item used was, “In a typical school week, estimate the number of hours you spend on the following for this school”. A sample response statement was, “Teaching students in the school”.

1. **Teachers’ Professional Collaboration**

In order to collect information about teachers’ cooperation and collaboration, teachers were asked to respond to a four-point Likert-type item, with the points ranging from 1 (never or almost never) to 4 (daily or almost daily). A sample item was, “How often do you have the following types of interactions with other teachers?” followed by a sample response statement like, “Discussion about how to teach a particular concept”.

1. **Teachers’ Professional Development**

The information about teachers’ professional development was collected through questionnaire items on the following areas: professional development programmes availed by teachers, teachers’ professional development needs, and barriers to teachers’ participation in professional development programmes. A sample item on the types of professional development programmes availed by teachers was, “In the past two years, have you participated in professional development in any of the following?” followed by a sample response statement, “Mathematics content”. The information about the teachers’ professional development needs was collected through a four-point Likert-type item, with the points ranging from 1 (no need at all) to 4 (high level of need). A sample item was, “Thinking of your own professional development needs, please indicate the extent to which you have such needs in each of the areas listed”, followed by a sample response statement, “Content and performance standards in my main subject”. The information about the barriers to teachers’ participation in professional development programmes was collected by asking teachers to respond to a questionnaire item like, “If you were not participating in any professional development programmes in the past years, which one of the following reasons best explains what prevented you from participating in professional development?” followed by a sample response statement, “Professional development was too expensive”.

1. **Teacher Appraisal**

The information about the priorities accorded by school principals to various areas of teacher appraisal was collected through a five-category Likert-type item of 16 response statements, with the categories ranging from 1 (do not know) to 5 (viewed with high importance). A sample item was, “In your opinion, how important were the following aspects viewed to be when you received appraisal and /or feedback from principal?” followed by a sample response statement, “Student test score”.

1. **Text Book Usage**

The information about how teachers’ used mathematics text books was collected through their responses to a dichotomous item. A sample item was, “Do you use a textbook in teaching mathematics to Grade 10 students?” followed by a sample response statement like, “As the primary basis for my lesson”.

1. **Teachers’ Readiness to Teach**

The teachers’ preparedness to teach Grade 10 mathematics was assessed through their responses to a three-point Liker-type item, with the points ranging from 1 (very ready) to 3 (not ready). A sample item was, “Considering your training and experience in both mathematics and content instruction, how ready do you feel you are to teach the following topics to your students?” A sample response statement was, “Matrices and network”.

1. **Mathematics Homework**

Teachers were asked to provide information about the frequency of homework, the duration of homework, the types of homework, and the kinds of feedback on homework that they provide to their students. A sample item on the frequency of homework was, “How often do you assign mathematics homework to your students?” with response statements varying from every lesson to some lessons. A sample item on the duration of homework was, “When you assign mathematics homework to your students, about how many minutes do you assign?” with the response statements varying from fewer than 15 minutes to more than 90 minutes. A three-point Likert-type item of three response statements, with the points ranging from 1 (always or almost always) to 3 (never or almost never) was used to measure the types of homework assigned by the teachers. A sample item was “How often do you assign the following kinds of homework to your students?” followed by a sample response statement “Doing problem/question sets”. Similarly, a three-point Likert-type item was used to measure the types of feedback given by the teachers on students’ homework. A sample item was, “How often do you do the following with the mathematics homework assignments?” followed by a sample response statement, “Use the homework to contribute towards students’ grades or marks”.

1. **Assessment Practices**

Teachers were asked to provide information about the frequency of mathematics tests and the types of questions that they used in the tests. A sample item on the frequency of mathematics test was, “How often do you give mathematics test or examination to your students? The response statements to the item ranged from about once a week to never. A three-point Likert-type item of three response statements, with the points ranging from 1 (always or almost always) to 3 (never or almost never), was used to collect information about the types of questions that the teachers used in mathematics tests. A sample item was, “How often do you include the following types of questions in your mathematics tests or examinations?” followed by a sample response statement, “Questions involving application of mathematical procedures”.

1. **Calculators and Computers**

Teachers were asked to provide information about how they used calculators and computers in their mathematics classes. A four-point Likert-type item of four response statements, with the points ranging form 1 (every or almost every lesson) to 4 (never), was used to assess the teachers’ use of calculators in the mathematics class. A sample item was, “How do students in your class use calculators for the following activities?” followed by a sample response statement “Check answers”. Similarly, a four-point Likert-type item of four response statements was used to assess the teachers’ use of computers in the mathematics class. A sample item was, “In your class, how often do you have students use a computer for the following activities?” followed by a response sample response statement “Discover mathematics principles and concepts”.

1. **Scale Indices and Construct Validation**
2. **Teachers’ Professional Cooperation**

Teachers’ professional collaboration was measured with a 4-category Likert-type item of four response statements, with the categories ranging from 1 (never or almost never) to 4 (daily or almost daily). A sample item was, “How often do you have the following types of interactions with other teachers?” followed by a sample response statement, “discussion about how to teach a particular concept”. However, the response category of “never or hardly ever” for some response statements was marked by only a few teachers, and the response category of “daily or almost daily” and “1-3 times per week” were not at all marked for some response statements. In order to perform a confirmatory factor analysis, the response categories of “Never or almost never” and “Daily or almost daily” were collapsed with the response categories of “2 or 3 times per month” and “1 to 3 times per week” respectively. The transformed item had a dichotomous response category of “2 to 3 times per month” and “1 to 3 times per week.”

A confirmatory factor analysis of the one-dimensional model of the teachers’ professional cooperation, using polychoric correlations with weighted least squares estimation method, showed a good model fit (RMSEA=0.00, NNFI=1.05, CFI=1.00, and RMR=0.70), indicating the unidimensionality of the response statements. The reliability of the item was 0.74. The regression coefficient of item 8a was not significant, with t-value being 1.61; while the regression coefficients of all other items were significant, with the smallest and the largest t-values being 2.88 and 4.67 respectively.

1. **Teachers’ Beliefs about Teaching**

Teachers’ beliefs about teaching were categorised into two broad teaching paradigms, namely, behaviourism and constructivism. These broad paradigms were measured with a Likert-type item, with response categories ranging from 1 (agree a lot) to 4 (disagree a lot). The item had 17 response statements, with eight response statements measuring the behaviourism and the rest measuring the constructivist belief. A sample item was, “To what extend do you agree or disagree with each of the following statements?” followed by a sample response statement “Mathematics should be learned as sets of algorithms or rules that cover all possibilities.”

A confirmatory factor analysis of the two-factor model of the teachers’ beliefs about teaching could not be performed by using all the 17 items simultaneously, due to insufficient sample size of only 60 teachers. Further, not all the response categories of all the items were used by the teachers, that is, some items behaved as dichotomous items. Therefore, the two underlying constructs of the teachers’ beliefs about teaching were mapped to their respective items, and separate reliability indices were calculated for each construct. Cronbach’s Alpha of both the constructs were weak (constructivism=0.36 and direct transmission=0.58). Therefore, the use of these constructs in interpreting teachers’ beliefs about teaching needs to be interpreted with caution in the chapter on data analysis, that is, Chapter 5 of this thesis. However, the items could be used as single items in assessing teachers’ beliefs about teaching.

1. **Classroom Teaching Practices**

Teachers’ classroom teaching practices were broadly categorised into the following categories of constructs: structured teaching, student-oriented teaching, enhanced activities, and the use of re-enforcements. A Likert-type item of 29 response statements were used to measure these constructs, with the response categories ranging from 1 (never or hardly ever) to 5 (in almost every lesson). 12 response statements measured the construct ‘structured teaching’, 10 response statements measured the construct ‘student-oriented teaching’, three response statements measured the construct ‘enhanced activities’, and four response statements measured the construct ‘the use of re-enforcements’.

Since the sample size of 60 teachers was not sufficient for performing a confirmatory factor analysis of a four-dimensional model of teachers’ classroom teaching practices, a confirmatory factor analysis on the item covariances with maximum likelihood estimation method was performed on each construct separately. First, the confirmatory factor analysis of the nine of the 12 items that measured the construct of structured teaching practice showed a moderate fit (RMSEA=0.09, NNFI=0.87, CFI=0.90, and RMR=0.14), confirming the unidimensionality of the items. The reliability of the nine items that measured the construct of structured teaching practice was 0.75. The regression coefficients of the nine items on their underlying constructs of structured teaching practice were all significant; with the smallest and the largest t-values being 2.19 and 5.25 respectively, indicating convergent validity. Second, the confirmatory analysis of the 10 items that measured the construct of student-oriented teaching practice showed a good model fit (RMSEA=.09, NNFI=0.91, CFI=0.91, and RMR=0.11), proving the unidimensionality of the items. The reliability of the items was 0.84. The regression coefficients of the 10 items on the construct of student-oriented teaching practice were significant; with the smallest and the largest t-values being 2.50 and 7.37 respectively, indicating convergent validity. Third, the confirmatory factor analysis of the three items that measured the construct of enhanced student activities showed a good model fit (RMSEA=0), confirming the unidimensionality of the items. The reliability of the three items was 0.73. The regression coefficients of the three items on their underlying construct were significant; with the smallest and the largest t-values being 3.80 and 5.71 respectively, confirming convergent validity. Fourth, the confirmatory factor analysis of the four items on the use of re-enforcements showed a moderate model fit (RMSEA=0.10, NNFI=0.93, CFI=0.98, and RMR=.09), providing evidence for the unidimensionality of the items. The reliability of the four items was 0.82. The regression coefficients of the four items were significant; with the smallest and the largest t-values being 3.99 and 13.23 respectively, indicating convergent validity.

1. **Factors Limiting Effective Teaching**

Teachers were asked to report on a five-point Likert-type item of six response statements, with the response categories ranging from 1 (not applicable) to 5 (a lot). A sample item was, “In your view, to what extent do the following limit how you teach mathematics effectively?” followed by a sample response statement, “Students with different academic abilities.”

A confirmatory factor analysis of the item covariances showed a good model fit (RMSEA=0.09, NNFI=0.92, CFI=0.95, and RMR=0.09), proving the unidimensionality of the item. The reliability of the item was 0.76. The regression coefficient of one of the six items (18b) was not significant, with t-value being 1.75. It was found that the item exhibited a large positive residual with another item (18a). Both the items were related to students’ ability. On removing item 18b, the fit of the model improved significantly (RMSEA=0.0, NNFI=1.07, CFI=0.10, and RMR=0.03). The reliability of the item with five response statements was 0.79. The regression coefficients of the five items were significant; with the smallest and the largest t-values being 2.62 and 7.32 respectively, providing evidence of convergent validity of the five response statements.

1. **Teachers’ Self-Efficacy**

Teachers’ self-efficacy was measured by asking teachers to respond to a 4-category Likert-type item of five response statements, with the categories ranging from 1 (strongly agree) to 4 (strongly disagree). A sample item was, “How far do you agree with the following?” followed by a sample response statement, “All in all, I am satisfied with my job.”

A confirmatory factor analysis of the item covariances showed a good model fit (RMSEA=0.0, NNFI=1.08, CFI=1.0, and RMR=0.01), indicating the unidimensionality of the five response statements. The reliability of the item was 0.74. The regression coefficients of the five response statements were significant; with the smallest and the largest t-values being 3.16 and 6.33 respectively, showing convergent validity.

1. **Teachers’ View of Classroom Climate**

Teachers’ view of classroom climate was measured by asking teachers to respond to a 4-category Likert-type item of five response statements, with the categories ranging from 1 (strongly disagree) to 4 (strongly agree). A sample item was, “How strongly do you agree or disagree with the following statements about your mathematics class?” followed by a sample response statement, “When the lesson begins, i have to wait quite a long time for students to pay attention”.

A confirmatory factor analysis showed the item covariances showed a good model fit (RMSEA=0.0, NNFI=1.05, CFI=1.00, and RMR=0.01), indicating the unidimensionality of the five response statements. The reliability of the item was 0.69. The regression coefficients of the five response statements were significant; with the smallest and the largest t-values being 2.76 and 5.81 respectively, showing convergent validity.

1. **Teachers’ View of School Climate**

Teachers’ view of school climate was measured by asking teachers to respond to a 4-category Likert-type item of four response statements, with the categories ranging from 1 (strongly disagree) to 4 (strongly agree). A sample item was, “How far do you agree with the following?” followed by a sample response statement, “In this school, teachers and students usually get on well with each other”.

A confirmatory factor analysis showed the item covariances showed a good model fit (RMSEA=0.08, NNFI=0.96, CFI=0.98, and RMR=0.01), indicating the unidimensionality of the four response statements. The reliability of the item was 0.84. The regression coefficients of the five response statements were significant; with the smallest and the largest t-values being 3.57 and 10.28 respectively, showing convergent validity.

1. **SCHOOL QUESTIONNAIRE**
2. **Simple Indices**
3. **School Size**

School principals were asked to provide information about the total number of boys and girls enrolled in their schools at the time of the research. The school size was defined in terms of the total number of boys and girls enrolled in the schools.

1. **School Funding Sources**

School principals were asked to provide the proportions of fund that their schools received from the following sources: the government, the student fees, and the benefactors or fund raising activities. Also, the principals were provided with the option of reporting the funding sources not included in the questionnaire item.

1. **Instructional Time**

School principals were asked to provide information about the number of instructional weeks available in the school year, the number of hours in the school week, and the number of instructional hours in the school week. The index of instructional time was calculated as the product of the instructional weeks available in the school year and the number of instructional hours in the school week.

1. **Availability of Computers**

Principals were asked to provide information about the number of computers available at their school. The index of the availability of computers was developed as a ratio of the number of computers to the number of students. Also, principals were asked to report the availability of local area network and the Internet service at their schools.

1. **Quantity of Teachers**

Principals were asked to report the number of both temporary and permanent mathematics teachers available in their schools. Also, principals were asked to report the teachers’ qualifications. The index of student to mathematics teacher ratio was developed by dividing the number of enrolled students by the total number of mathematics teachers.

1. **School Admittance Practices**

School admittance practices were assessed with a four-category rating item of six response statements, with the categories ranging from 1 (pre-requisite) to 4 (not considered). A sample item was, “How much consideration is given to the following factors when students are admitted to your school?” followed by a sample response statement, “Residence in a particular area.” The index of school admittance practices was constructed by categorizing schools into any one of the four response categories.

1. **Assessment Practices**

School-based assessment practices were assessed with a five-category rating item of five response statements, with the categories ranging from 1 (never) to 5 (more than once a month). A sample item was, “Generally, in your school, how often are students assessed using the following assessments?” followed by a sample response statement, “standardised tests.” In addition, school principals were asked to answer a dichotomous item of eight response statements that described different functions of assessments. A sample item was, “In your school, are assessments of students used for any of the following purpose?” followed by a sample response statement, “To inform parents about their child’s progress.”

1. **Ability Grouping**

School principals were asked to report if they offered student ability-based instructions at their schools. A three-point rating item of two response statements, the points ranging from 1 (for all classes) to 3 (not for any classes), was used in collecting information about the practices related to student ability grouping. A sample item was, “Schools sometimes organise instruction differently for students with different abilities and interests in mathematics. Which of the following options describe what your school does for students in mathematics classes?” followed by a sample response statement, “Students are grouped by ability within their mathematics classes.”

1. **Mathematics Enrichment Activities**

School principals were asked to provide information about the types of mathematics enrichment activities conducted in their schools. A dichotomous item of five response statements was used to collect information about the types of mathematics enrichment activities. A sample item was, “In your school, do any of the following activities to promote engagement with mathematics occur?” followed by a sample response statement, “Remedial mathematics.”

1. **School Autonomy**

The information about school autonomy was collected by asking principals about who is responsible for school resources and curriculum. A sample item on resource autonomy was, “In your school, who has the main responsibility for formulating the school budget?” followed by a set of sample response options of “Not a main responsibility of the school”; “School Management Board”; “School Head”, “Subject Head”; and “Teachers”. Similarly, a sample item on school curriculum autonomy was, “In your school, who has the main responsibility for choosing which text books are used?” followed by a set of sample response options of “Not a main responsibility of the school”; “School Management Board”; “School Head”, “Subject Head”; and “Teachers”. Also, school principals were asked to provide information about who influenced them in making decision on staffing, budgeting, instructional content, and assessment practices. A sample item on who influences the principals’ decision making power was, “In your school, which of the following exert a direct influence on decision making about staffing, budgeting, instructional content, and assessment practices?” followed by a sample response statement, “national education authorities.”

1. **Monitoring Strategies**

School principals were asked to provide information about different monitoring strategies that they used in their schools to monitor mathematics teachers. A dichotomous item of four response statements was used to collect information about the monitoring strategies used by the principals. A sample item was, “During the last year, have any of the following been used to monitor the practice of mathematics teachers at your school?” followed by a sample response option, “Tests or assessments of student achievement.”

1. **Scale Indices and Construct Validation**
2. **Quality of School Resources**

School resources were grouped into the following three categories: physical infrastructure, educational resources, and teacher shortage. The quality of schools’ physical infrastructural resources was measured with a four-category Likert-type item of three response statements, with the categories ranging from 1 (not at all) to 4 (a lot). A sample item was, “Is your school’s capacity to provide instruction hindered by a shortage or inadequacy of any of the following?” followed by a sample response option, “school buildings and grounds.” The quality of schools’ educational resources was measured with a four-category Likert-type item of four response statements, with the categories ranging from 1 (not at all) to 4 (a lot). A sample item was, “Is your school’s capacity to provide instruction hindered by a shortage or inadequacy of any of the following?” followed by a sample response option, “availability of instructional materials.” Similarly, the schools’ teacher shortage problem was measured by a four-category Likert-type item of three response statements, with the categories ranging from 1 (not at all) to 4 (a lot). A sample item was, “Is your school’s capacity to provide instruction hindered by a shortage or inadequacy of any of the following?” followed by a sample response option, “availability of qualified mathematics teachers.”

The quality of school resources was a three-dimensional construct, with the constructs being the physical infrastructure, educational resources, and teacher shortage. However, it was not feasible to perform a confirmatory factor analysis of the three-dimensional model because of the small sample size of 60 schools. Therefore, a separate confirmatory factor analysis was performed on each dimension. First, the confirmatory factor analysis of the item covariances with maximum likelihood estimation method on the measures of the schools’ physical infrastructural resources showed a good model fit (RMSEA=0.00), indicating the unidimensionality of the measures. The regression coefficients of the measures on their underlying construct were all significant, with the smallest and the largest t-values being 3.59 and 5.33 respectively, showing convergent validity. The composite reliability of the measures was 0.74. Second, the confirmatory factor analysis on the measures of the schools’ educational resources showed a good model fit (RMSEA=0.00, NNFI=1.01, CFI=1.00, and RMR=0.03), indicating the unidimensionality of the measures. The regression coefficients of the measures on their underlying construct were all significant, with the smallest and the largest t-values being 2.34 and 8.61 respectively, indicating convergent validity. The composite reliability of the measures was 0.77. Third, the confirmatory factor analysis of the item covariances with maximum likelihood estimation method on the measures of the schools’ teacher shortage showed a good model fit (RMSEA=0.00, NNFI=1.11, CFI=1.00, and RMR=0.03), showing the unidimensionality of the measures. The regression coefficients of the measures on their underlying construct were all significant, with the smallest and the largest t-values being 3.26 and 4.85 respectively, confirming convergent validity. The composite reliability of the measures was 0.71.

1. **Teacher Morale**

School principals were asked to respond to a 4-category Likert-type item of four response statements, with the categories ranging from 1 (strongly agree) to 4 (strongly disagree). A sample item was, “Think about the teachers in your school. How much do you agree with the following statements?” followed by a sample response statement, “The morale of teachers in this school is high.” Of the four response statements, two had only three categories,1 to 3, marked by the school principals, and the other two had only two categories, 1 and 2, marked by the school principals. Since the response statements with only two categories marked by the school principals failed to discriminate the school principals who agreed from those who did not agree with response statements, they were excluded from index of teacher morale. However, the two response statements with only two response categories marked would be analysed separately to retain the information. Therefore, the index of the teacher morale was developed with a 3-category Likert-type item of two response statements, with categories ranging from 1 (strongly agree) to 3 (disagree).

The confirmatory factor analysis on the item covariances with maximum likelihood estimation method of the one-dimensional model of the teacher morale showed a good model fit (RMSEA=0.00), indicating the unidimensionality of the measures. The regression coefficients of the measures on their underlying construct, that is, the teacher morale were all significant, with the t-values being 5.83 and 10.86. The composite reliability of the measures was 0.84.

1. **Teacher Consensus**

School principals were asked to respond to three 4-category Likert-type items of three response statements each, with the categories ranging from 1 (strongly agree) to 4 (strongly disagree). All positively worded response statements had the third and the fourth response categories blank, that is, school principals either strongly agreed or agreed. On the other hand, the negatively worded response statements had the first and the fourth response categories blank, that is, the school principals either agreed or disagreed. The index of teacher consensus was developed with the negatively worded response statements. The response categories “agree” and “disagree” were recoded to 1 and 2 respectively, so that high values indicated more characteristics of teacher consensus. A sample item was, “How much do you agree with these statements about innovation in your school?” followed by a sample response statement, “There are frequent disagreements between innovative and traditional mathematics teachers.”

The confirmatory factor analysis of the item correlations with robust maximum likelihood estimation method showed a good model fit (RMSEA=0.00), indicating the unidimensionality of the measures. The regression coefficients of the measures on the underlying construct were all significant, with the minimum and the maximum t-values being 2 and 3.57 respectively, indicating convergent validity. The composite reliability coefficient of the measures was 0.72.

1. **School Climate**

School principals were asked to report their perceptions of how the factors related to teachers and students affect school climate. A 4-category Likert-type item of 13 response statements, with the categories ranging from 1 (not at all) and 4 (a lot) asked the school principals about the extent to which the learning of students was hindered by the factors related to teachers and students. Since very few school principals had responded in the fourth response category, the fourth response category was collapsed with the third response category. Seven of the 13 response statements were on the teacher-related factors and the remaining six response statements were on the student-related factors. A sample item was, “In your school, to what extent is the learning of students hindered by:” followed by a sample response statement of “teachers’ low expectations of students?”

The confirmatory factor analysis of the item covariances of the teacher-related factor of the school climate with maximum likelihood estimation method showed a good model fit (RMSEA=0.00, NNFI=1.02, CFI=1.00, and RMR=0.03), indicating the unidimensionality of the response statements. The regression coefficients of the response statements on their underlying construct were all significant, with the smallest and the largest t-values being 3.13 and 7.60 respectively, indicating the convergent validity of the response statements. The composite reliability of the response statements was 0.79. Similarly, the confirmatory factor analysis of the item covariances of the student-related factor of the school climate with maximum likelihood estimation method showed a moderate model fit (RMSEA=0.16, NNFI=0.80, CFI=0.88, and RMR=0.04), indicating the unidimensionality of the response statements. The regression coefficients of the response statements were all significant, with the smallest and the largest t-values being 3.18 and 5.05 respectively, indicating the convergent validity of the response statements. The composite reliability of the response statements was 0.76.

1. **School Leadership**

School principals were asked to provide information about how frequently they performed 10 instructional leadership functions (e.g., developing school goals, providing incentives to staff) in their schools. The items were adapted from PIMRS ([Hallinger, 1990](#_ENREF_3)). Each leadership function was measured with a 5-point rating item of five response statements, with the points ranging from 1 (almost never) to 5 (almost always). A sample item on developing school goal was, “To what extent do you do the following tasks?” followed by a sample response statement, “develop a focussed set of annual school-wide goals.” Very few principals had responded in the rating cells 1 and 2. Therefore, the rating points 1 and 2 were collapsed for the purpose of performing construct validation. The following paragraphs describe the construct validation procedures used in validating each one of the 10 leadership aspects.

The confirmatory factor analysis of the item covariances of the goal setting aspect of instructional leadership showed a moderate model fit (RMSEA=0.12, NNFI=0.87, CFI=0.93, and RMR=0.03), indicating the unidimensionality of the response statements. The regression coefficients of the response statements on their underlying construct were all significant, with the smallest and the largest t-values being 2.72 and 6.07 respectively, showing the convergent validity of the response statements. The composite reliability of the response statements was 0.74.

The confirmatory factor analysis of the item covariances of the goal communication aspect of the leadership with maximum likelihood estimation method showed a good model fit (RMSEA=0.11, NNFI=0.97, CFI=0.97, and RMR=0.03), indicating the unidimensionality of the response statements. The regression coefficients of the response statements on their underlying construct were all significant, with the smallest and the largest t-values being 4.44 and 8.39 respectively, indicating the convergent validity of the response statements. The composite reliability of the response statements was 0.81.

The confirmatory factor analysis of the item covariances of supervision and evaluation aspect of the leadership with maximum likelihood estimation showed a good model fit (RMSEA=0.12, NNFI=0.90, CFI=0.95, and RMR=0.05), indicating the unidimensionality of the response statements. The regression coefficient of one of the response statements was insignificant, with t-value of 2.16, failing the convergent validity test. However, the other four response statements loaded significantly on their underlying construct, with the smallest and the largest t-values of 3.09 and 8.75 respectively, indicating the convergent validity. The composite reliability of the response statements was 0.79.

The confirmatory factor analysis of the item covariances of curriculum coordination aspect of the leadership with maximum likelihood estimation method resulted in a poor model fit. The inspection of the residuals showed that two response statements had significant correlation, and also the regression coefficient of one of the two response statements was insignificant. The confirmatory factor analysis of the item covariances on the remaining response statements showed a good model fit (RMSEA=0.00, NNFI=1.00, CFI=1.00, and RMR=0.03), showing the unidimensionality of the response statements. The regression coefficients of the response statements on their underlying construct were significant, with the smallest and the largest t-values of 3.09 and 7.18 respectively, indicating the convergent validity. The composite reliability of the response statements was 0.79.

The confirmatory factor analysis of the item covariances of monitoring aspect of the leadership with maximum likelihood estimation method showed a good model fit (RMSEA=0.11, NNFI=0.93, CFI=0.96, and RMR=0.037), indicating the unidimensionality of the response statements. The regression coefficients of the response statements on their underlying constructs were all significant, with the smallest and the largest t-values of 4.10 and 6.28 respectively, showing the convergent validity of the response statements. The composite reliability of the response statements was 0.80.

The confirmatory factor analysis of the item covariances of the leadership aspect of safeguarding instructional time, with maximum likelihood estimation method, showed a good model fit (RMSEA=0.04, NNFI=0.99, CFI=0.99, and RMR=0.04), indicating the unidimensionality of the response statements. The regression coefficients of the response statements on their underlying construct were all significant, with the smallest and the largest t-values of 3.17 and 5.18 respectively, indicating the convergent validity of the response statements. The composite reliability of the response statements was 0.73.

The confirmatory factor analysis of the item covariances of the visibility aspect of the leadership, with maximum likelihood estimation method, showed a moderate model fit (RMSEA=0.16, NFFI=0.82, CFI=0.91, and RMR=0.05), indicating the unidimensionality of the response statements. The regression coefficient of one of the response statements was insignificant, with the t-value of 2.07. All other response statements loaded significantly on their underlying construct, with the smallest and the largest t-values of 3.33 and 7.20, confirming the convergent validity. The composite reliability of response statements was 0.73.

The confirmatory factor analysis of the item covariances of providing teacher incentives aspect of the leadership, with maximum likelihood estimation method, showed a good model fit (RMSEA=0.02, NNFI=0.99, CFI=1.00. RMR=0.03), indicating the unidimensionality of the response statements. The regression coefficients of the response statements on their underlying construct were all significant, with the smallest and the largest t-values of 4.96 and 6.10 respectively, indicating the convergent validity of the response statements. The composite reliability of the response statements was 0.82.

The confirmatory factor analysis of the item covariances of promoting professional development aspect of the leadership, with maximum likelihood estimation method, showed a moderate model fit (RMSEA=0.13, NNFI=0.93, CFI=0.96, and RMR=0.04), indicating the unidimensionality of the response statements. The regression coefficients of the response statements on their underlying construct were significant, with the smallest and the largest t-values of 5.57 and 7.87 respectively, showing the convergent validity of the response statements. The composite reliability of the response statements was 0.84.

The confirmatory factor analysis of the item covariances of providing incentives for student learning aspect of the leadership, with maximum likelihood estimation method, showed a moderate model fit (RMSEA=0.15, NNFI=0.85, CFI=0.93, and RMR=0.05), indicating the unidimensionality of the response statements. The regression coefficient of one of the response statements was insignificant, with the t-value of 2.14. The regression coefficients of all other response statements were significant, with the smallest and the largest t-values of 3.48 to 7.07 respectively, indicating the convergent validity. The composite reliability of the response statements was 0.73.

**References**

Ganzeboom, H. B. G., Graaf, P. M. D., Treiman, D. J., & Leeuw, J. D. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research, 21*(1), 1-56.

Ganzeboom, H. B. G., & Treiman, D. J. (1996). Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research, 25*(3), 201-239.

Hallinger, P. (1990). Principal Instructional Management Rating Scale: Principal Form 2.0. Sarasota, FL: Leading Development Associates.

International Labour Office. (1990). *ISCO-88: International Standard Classification of Occupations*.

OECD. (1999). *Classifying Educational Programmes: Manual for ISCED-97 Implementation in OECD Countries*. Retrieved from <http://www.oecd.org/dataoecd/41/42/1841854.pdf>.

OECD. (2005). *PISA 2003 Technical Report*. Paris: OECD.

OECD. (2009). *PISA 2006 Technical Report*. Paris: OECD.

UNESCO. (1997). ISCED 1997, International Standard Classification of Education Available from [www.uis.unesco.org/TEMPLATE/pdf/isced/ISCED\_A.pdf](http://www.uis.unesco.org/TEMPLATE/pdf/isced/ISCED_A.pdf)